

## **NIEHS contributes to big NIH investment in biomedical research data**

By Joe Balintfy

As part of wide-ranging grants announced Oct. 10 by NIH, NIEHS is helping develop new strategies to analyze and make good use of the explosion in complex biomedical data sets, often referred to as Big Data. NIEHS involvement in the NIH Big Data to Knowledge ( [BD2K](http://bd2k.nih.gov/index.html#sthash.EtnLPva4.dpbs) (<http://bd2k.nih.gov/index.html#sthash.EtnLPva4.dpbs>) ) investment spans all areas of emphasis, including training, research centers, and data discovery.

“NIEHS has had a role in developing BD2K initiatives through our participation on numerous workgroups over the past 18 months,” said Allen Dearry, Ph.D., director of the NIEHS Office of Scientific Information Management and NIEHS representative to the BD2K Executive Committee.

### **Big data centers**

“We are managing the career development portion of this nearly \$32 million investment,” said Gwen Collman, Ph.D., director of the NIEHS Division of Extramural Research and Training. “We are also participating in the grants establishing Centers of Excellence for Big Data Computing.”

#### The 11 centers

(<http://bd2k.nih.gov/FY14/COE/COE.html#sthash.unlq1Shw.dpbs>)

will develop innovative approaches, methods, software, tools, and other resources to enhance access to data and the ability to make new discoveries using it, according to an NIH [press release](http://www.nih.gov/news/health/oct2014/od-09.htm) (<http://www.nih.gov/news/health/oct2014/od-09.htm>)

. Development efforts will focus on specific research questions, yet the centers’ output is expected to be more generally relevant to aspects of big data science, such as data integration and use, analysis of genomic data, and managing data from electronic health records.

### **Data science training and workforce development**

The Training and Career Development in Biomedical Big Data [awards](http://bd2k.nih.gov/FY14/KO1/KO1.html#sthash.johNY9A3.dpbs)

(<http://bd2k.nih.gov/FY14/KO1/KO1.html#sthash.johNY9A3.dpbs>)

support education and training of researchers who will specialize in data science fields, as well as those whose work may require expertise in the use or generation of large amounts of data and data resources.

“Some training projects include brain imaging studies, identifying healthy behaviors that may reduce risk of heart disease and diabetes, and harnessing the largely untapped potential of social media data to capture social and cultural processes with potential health impact,” said NIEHS Program Director Carol Shreffler, Ph.D.

Teams at large research universities and academic medical centers may have bioinformatics and data support, but individual scientists in the biomedical research community may not, and many have not been trained to access and analyze large data sets.

### **Data discovery**

NIH also launched an effort to pilot a [Data Discovery Index](http://bd2k.nih.gov/FY14/DDICC/DDICC.html#sthash.2pUPInJF.dpbs)

(<http://bd2k.nih.gov/FY14/DDICC/DDICC.html#sthash.2pUPInJF.dpbs>)

(DDI), which will catalyze the discoverability, accessibility, and citation standards for biomedical big data. “Currently, there is no easy query or search infrastructure that can help identify the presence and availability of relevant data sets,” said Becky Boyles, a member of the NIH team overseeing this project and an NIEHS data scientist.

Boyles explained that data are found in an increasing number and variety of different repositories or web sites, when they are available at all. “The intent of establishing a DDI is to help researchers find, reuse, and cite relevant publicly available datasets related to their scientific question of interest,” she said.



*“We have been instrumental in ensuring inclusion of environmental and exposure data within the diverse types of big data addressed by BD2K,” said Dearry. (Photo courtesy of Steve McCaw)*



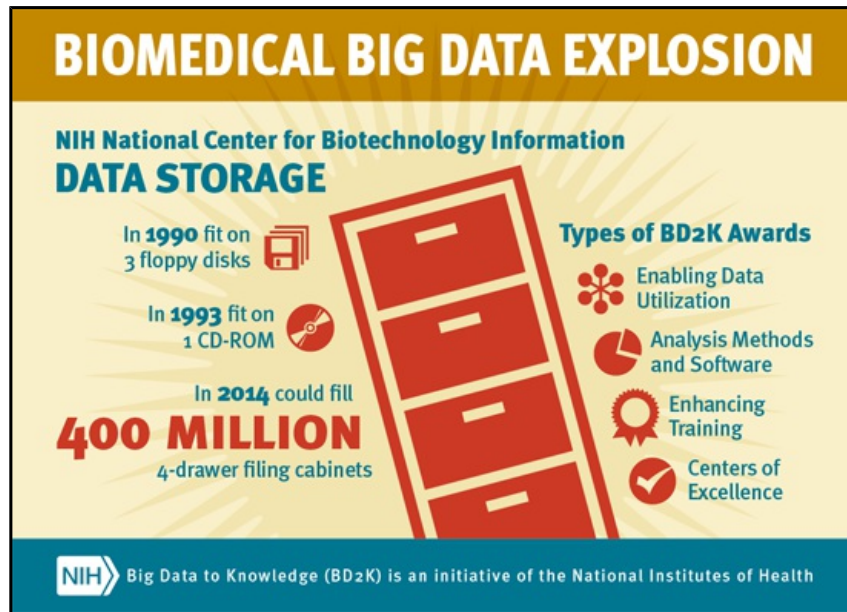
*“The training programs span a wide range of big data research areas,” said Shreffler. (Photo courtesy of Steve McCaw)*

## Big data is a big deal

Studies generating billions of data points continue to proliferate. For example, environmental studies gather data on multiple exposures and health outcomes, epidemiological studies examine thousands of participants for health and disease patterns, large disease-oriented efforts seek the genomic underpinnings of illnesses, and other projects look to identify all functional elements in the human genome.

“Mammoth data sets are emerging at an accelerated pace in today’s biomedical research,” said NIH Director Francis S. Collins, M.D., Ph.D. “The potential of these data, when used effectively, is quite astounding.”

(Joe Balintfy is a public affairs specialist in the NIEHS Office of Communications and Public Liaison.)



*Big data comes along with challenges such as locating data, lack of data standards, and lack of tools to access and analyze data sets.*

---

The Environmental Factor is produced monthly by the [National Institute of Environmental Health Sciences \(NIEHS\)](http://www.niehs.nih.gov/)

(<http://www.niehs.nih.gov/>)

, Office of Communications and Public Liaison. The content is not copyrighted, and it can be reprinted without permission. If you use parts of Environmental Factor in your publication, we ask that you provide us with a copy for our records. We welcome your [comments and suggestions](#). ([bruskec@niehs.nih.gov](mailto:bruskec@niehs.nih.gov))

This page URL: NIEHS website: <http://www.niehs.nih.gov/>

Email the Web Manager at [webmanager@niehs.nih.gov](mailto:webmanager@niehs.nih.gov)